

## Task Title: Fraud Detection in Credit Card Transactions

### Scenario:

XSPACE TECHNOLOGIES Pvt Ltd is partnering with a major financial institution that is facing challenges with detecting fraudulent credit card transactions. With the increasing volume of transactions, it has become difficult for traditional methods to accurately identify fraud in real-time. The company has provided a dataset containing transaction details, and your task is to build a machine learning model that can predict whether a transaction is fraudulent.

Your task is to analyze the data, build a model to detect fraudulent transactions, and provide recommendations on how the financial institution can improve its fraud detection system.

### Objective:

1. Understand the patterns that differentiate fraudulent transactions from legitimate ones.
2. Build a predictive model that accurately detects fraud.
3. Provide insights and recommendations for improving the fraud detection process.

### Dataset:

You are provided with a CSV file containing the following columns:

- **Transaction ID:** Unique identifier for each transaction.
- **Customer ID:** Unique identifier for each customer.
- **Transaction Date:** The date and time when the transaction occurred.
- **Transaction Amount:** The amount of money involved in the transaction.
- **Merchant:** The name of the merchant where the transaction took place.
- **Location:** The location of the transaction.
- **Transaction Type:** The type of transaction (e.g., online purchase, in-store purchase, ATM withdrawal).
- **Card Type:** The type of credit card used (e.g., Visa, MasterCard).
- **Is Fraudulent:** A binary indicator (Yes/No) of whether the transaction was fraudulent.

### Tasks:

1. **Data Understanding and Preprocessing:**
  - Perform data exploration to understand the distribution of fraudulent vs. legitimate transactions.
  - Handle any missing or inconsistent data.
  - Convert categorical variables into numerical representations where necessary (e.g., Merchant, Location, Transaction Type).
  - Balance the dataset if it is highly imbalanced (e.g., using techniques like SMOTE or undersampling).
2. **Feature Engineering:**
  - Create new features that might help in fraud detection, such as the frequency of transactions by a customer, the average transaction amount, or the time of day of transactions.
  - Explore interactions between different features to improve model performance.
3. **Model Building:**
  - Split the dataset into training and testing sets.
  - Experiment with different machine learning algorithms, such as Logistic Regression, Random Forest, or Gradient Boosting, to build a fraud detection model.
  - Tune the model for optimal performance using techniques like cross-validation and grid search.
  - Evaluate the model using metrics like accuracy, precision, recall, F1-score, and especially the area under the ROC curve (AUC-ROC) for imbalanced datasets.
4. **Fraud Detection System Design:**
  - Based on your findings, design a real-time fraud detection system that could be implemented by the financial institution.
  - Suggest methods for continuous learning and adaptation of the model as fraud patterns evolve.
  - Provide recommendations on how to integrate this system with the existing transaction processing infrastructure.
5. **Report Generation:**
  - Compile a comprehensive report that includes the data analysis, model development, evaluation results, and recommendations.
  - Include visualizations such as confusion matrices, feature importance charts, and AUC-ROC curves to illustrate model performance.

### Tools Required:

- Python (Pandas for data manipulation, Scikit-learn for machine learning, Matplotlib/Seaborn for visualization)
- Jupyter Notebook or Google Colab for analysis and model development

### Expected Outcome:

- A well-preprocessed and engineered dataset ready for modeling.
- A fraud detection model with strong performance metrics.
- A conceptual design for a real-time fraud detection system.
- A detailed report with insights, model evaluation, and recommendations.

## Submission Guidelines:

- Submit your code, dataset, and report via a GitHub repository.
- Ensure that your notebook is well-documented with comments explaining each step.
- The deadline for this task is **25-September-2024**.

## Evaluation Criteria:

- **Data Preprocessing and Feature Engineering:** Effectiveness in cleaning and enhancing the dataset.
- **Model Performance:** Accuracy and robustness of the fraud detection model.
- **System Design:** Creativity and practicality in designing a real-time fraud detection system.
- **Report Quality:** Clarity of insights, visualizations, and recommendations provided.